Exploiting the Cell BE Architecture on Roadrunner for Long-time Atomistic Dynamics

Danny Perez, Arthur F. Voter, T-12; Sriram Swaminarayan, CCS-2

The last few years have seen the demise of single-core superscalar microprocessor architectures in favor of multicore processors. With physical and manufacturing constraints and power density requirements inhibiting further increases in clock frequency, the industry is now turning to heterogeneous architectures such as multiple different cores and/or specialized vector co-processing units to deliver increased performance. One such next-generation architecture is IBM's Cell Broadband Engine (Cell BE) [1]. A Cell BE is composed of a general-purpose Power PC element (the PPE) to which are connected eight synergistic processing elements (SPEs), which are vector processing units. Taking full advantage of the SPEs can theoretically deliver an impressive 204 GFlop/s in single precision or 102 GFlop/s in double precision. The prospect of such stunning performance has generated a lot of excitement around this new architecture, culminating in the design of the upcoming LANL supercomputer, Roadrunner.

This paradigm shift in processor design poses serious challenges to scientific programmers. Indeed, these potential gains can only be realized at the price of extensive redesigns of existing applications. High-performance applications will now need to manage fine-grained, intra-processor parallelization over a heterogeneous collection of processing units on top of the usual coarser-grained inter-processor parallelization, the former now becoming increasingly performance-critical. New strategies are thus needed if the promises of next-generation supercomputers like Roadrunner are to be achieved.

In this communication, we present one such strategy aimed at achieving high performance in atomistic molecular-dynamics simulations on Cell BE-based supercomputers. We find that a time-based parallelization on systems of modest size using the parallel replica method can outperform the conventional space-based parallelization of large systems through a more efficient use of the SPEs. We demonstrate this using benchmarks on a single Cell BE processor and deduce the expected performance on Roadrunner.

The molecular-dynamics simulation method (MD) is one of the cornerstones on which our understanding of the atomic-scale behavior of materials is built. Traditionally, good performance of large-scale simulations has been tied to the minimization of interprocessor communication. This is typically accomplished by assigning distinct regions of space to different processors. Communication is then required to inform each processor of the current state of the atoms in neighboring regions of space. Maximal performance is achieved when the volume-to-surface ratio of each region is high, keeping the computation-to-communication ratio high. This strategy is efficient for simulating very large systems over relatively short time scales, making it ideal to study problems such as shock wave propagation [2].

An alternative strategy is to exploit a time-based decomposition of the problem using the parallel-replica dynamics method [3]. In this method, developed here at LANL, a complete replica of the system resides on each processor. For dynamics dominated by infrequent transitions between different configurations, one can show that the time required for a single replica to observe a transition is statistically equivalent to the sum of times taken by a number of independent replicas before any one of them observes a transition. One can thus obtain a nearly linear speed-up when increasing the number of processors. Since the replicas are independent, communication is only necessary to transmit the new state of the system each time a transition occurs. As such, communication is minimized for small systems where transitions are infrequent. This method is thus ideally suited to study slow processes, such as defect diffusion and annealing, over very long time scales [4].

Since these two simulation approaches are aimed at different regimes (size scale vs time scale), a given problem typically will be well suited to one and not the other. However, viewed purely from the perspective of efficiency, the parallel replica strategy offers distinct advantages on the Cell BE. The overall efficiency of simulations on Roadrunner-type supercomputers strongly depends on the efficient use of the SPEs. The challenge is formidable since the local memory on the SPEs [the so-called Local Store (LS)] is limited to 256 kB for both instructions and data. Thus, at most a few thousand atoms can be stored in the LS, the rest remaining in the Cell BE's main memory (which is 8 GB on Roadrunner). This implies that if the system assigned to each Cell BE is larger than this, atoms will need to be streamed through the LS of the SPEs a few thousand at a time to keep them continuously fed. Further, neighbor list information can not be cached in the LS and has to be recalculated during each force computation. The resulting overhead reduces the overall efficiency of the computation. The crucial point is that the constraints placed on the space-decomposition scheme by maximizing the coarse-grained volume-to-surface ratio (i.e., minimizing the communication beween Cell BEs) reduces the efficiency of the computation of the SPEs since systems will not fit in LS. However, this situation should be ideal for the parallel replica method.

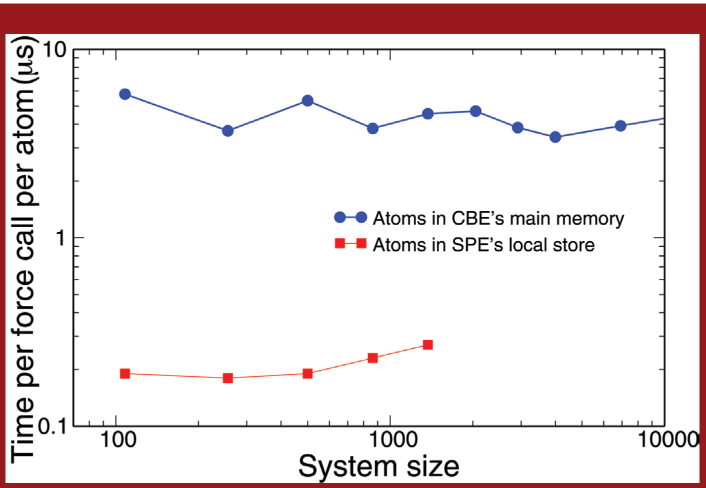

Fig. 1. Time per force computation per atom as one Cell BE processor for two storage strategies.

To test this hypothesis, we implemented force computations (the most time-consuming task in atomistic simulations) corresponding to the two aforementioned strategies on the Cell BE architecture using the embedded atom method (EAM) potential of Voter[5]. The two implementations essentially differ by the storage scheme of the atomic coordinates: in the first (corresponding to large-scale space-decomposition MD), they are stored in the Cell BE's main memory and streamed through the SPE's LS, while in the second (corresponding to the parallel replica method) the entire system is resident in each SPE's LS. The respective timings are reported in Fig. 1, where the time required to compute the force acting on a single atom using a single Cell BE processor is presented as a function of the system size.

The results clearly demonstrate the advantage of storing the system in the SPE's LS. Indeed, the time per force call per atom is around 0.2 μs in this case, compared with about 4 μs for storage in main memory, an increase in speed by a factor of 20. While precise values can be slightly affected by other parameters of the calculation, our timings clearly show that large performance gains can be achieved by dedicating each Cell BE to a system of small size. As mentioned above, this gain will be offset by communication costs if used in conjunction with a space decomposition strategy but is a perfect fit to the parallel replica method.

Using these results we can extrapolate the performance of a parallel replica code on Roadrunner. As an example, we recently used the parallel-replica method to study vacancy void dynamics in fcc metals, and discovered a surprising propensity for direct collapse to a stacking fault tetrahedron [4] (see *Direct Transformation of Vacancy Voids to Stacking Fault Tetrahedral,* pg. 160 in this book). If a similar study of post-collision annealing of radiation-produced defects was to be carried out on Roadrunner using 2000 atoms, it should be possible to simulate the evolution of the system at a rate of up to 1 *ms*/day of computation. Further, the parallel replica method can be combined with other accelerated dynamics methods, most notably the hyperdynamics method [6], to push further out (by anywhere from a factor of 10 to a factor of 10^5) the timescales that can be reached using atomistic simulations. Thus, Roadrunner will enable the parallel replica method to access experimentally relevant timescales for slow processes like aging or defect annealing studies, while retaining a fully atomistic description of the system.

In conclusion, we have shown that considerable performance gains can be obtained for atomistic simulations on the Cell BE architecture if each Cell BE is dedicated to a system of a few thousand atoms, a regime ideally suited for the the parallel replica method. Further, this strategy could also be used to design cheap, commodity-based computers (using Sony's PlayStation 3 game consoles, for example) geared toward long-time molecular-dynamics simulations. Our approach opens the door to efficient long-time simulations on next-generation supercomputers, nicely complementing the existing capabilities of large-scale molecular-dynamics and thus considerably extending the range of applicability of atomistic simulations.

For more information contact Danny Perez at danny_perez@lanl.gov.

[1] *Cell Broadband Engine Architecture,* IBM Systems and Technology Group, 330 Hopewell Junction, NY 12533-6351,1st ed. (2007).
[2] K. Kadau, et.al, *Phys. Rev. Lett.* **98,** 135701 (2007).
[3] A.F. Voter, *Phys. Rev. B* **57,** R13985(1998).
[4] B.P. Uberuaga, et al, *Phys. Rev. Lett.* **99,** 135501 (2007).
[5] A.F. Voter, Los Alamos National Laboratory Technical Report LA-UR-93-3901, Los Alamos National Laboratory(1993).
[6] A.F. Voter, *Phys. Rev. Lett.* **78,** 3908 (1997).

Funding Acknowledgments
- Department of Energy, Office of Science, Office of Basic Energy Sciences
- Los Alamos National Laboratory Directed Research and Development Program